

AD No. 19184  
ASTIA FILE COPY

Contract N8 onr - 66213 with the Office of Naval Research  
Project NR153-148: Performance Examinations for the Training  
and Selection of Scientific Personnel.

PERFORMANCE TESTS IN PHYSICS

Haym Kruglak

PHYSICS DEPARTMENT  
UNIVERSITY OF MINNESOTA

## Performance Tests in Physics

Haym Kruglak

University of Minnesota

### Abstract

The investigation extended previous research on performance tests in elementary college physics. The effect of high school physics on college physics achievement was explored. Correlation coefficients between performance tests and other measures of aptitude and achievement were calculated. The dependence of the test reliability on the number of test items and subscores was investigated.

It was found that:

1. High school physics background probably contributed favorably to achievement in college physics as measured by conventional examinations, but the effect on performance tests scores was doubtful.
2. The correlations between the performance test, ACE and high school rank were not significantly different from zero.
3. The correlations between the performance test and conventional achievement criteria in college physics were positive, significant, but very low.
4. The Hoyt reliability coefficient of a performance test could be increased to an acceptably high value by entering several subscores for each item.

### Performance Tests in Physics\*

This study is an extension of the previously reported research on performance tests in physics (3, 4). During the spring quarter 1951-52 additional samples were drawn from the general courses in physics at the University of Minnesota. The effect of high school physics on achievement in college physics was further investi-

---

\* Under contract N8onr-66213 with the Office of Naval Research, Project NR153-148

gated; a number of changes were made in the performance tests; some of the original findings were checked; and the test reliability was explored more thoroughly.

### Effect of High School Physics

The test results for Physics 7, fall quarter 1951, showed no significant differences on three achievement measures between students who had had high school physics and those who had none (3). The analysis was repeated for a comparable sample drawn from two courses offered during spring quarter 1952. It is apparent from Table 1 that the conclusions from the second investigation appear to contradict the original findings.

There was a significant difference in means for the spring quarter group for the final course grade, Cooperative Physics Test for College Students and the laboratory performance test in favor of those who had had high school physics. This difference between the two groups could hardly be attributed to the fact that the fall quarter course contained largely sophomores, whereas the spring quarter sample was composed primarily of third quarter freshmen. Both groups - freshmen and sophomores - were adjudged to be random samples from a normal population on the basis of three initial measures: High School Percentile Rank, ACE, and Cooperative Test in High School Physics.

When the subsamples from both quarters were pooled the means were still significantly different for the final grade and the cooperative test. Since four of the nine items on the performance test were identical and since the means for the two quarters differed by only one point, it was thought reasonable to pool all the students on this test and compare those who had had high school physics with those without it. It was found that there were no significant differences in the variances or the means. However, the number of students entering engineering physics at the University of Minnesota without high school physics is relatively small: 10 to 15 per

Table 1

## Achievement of Students with and without High School Physics

Physics 7, University of Minnesota, 1951-52

Final Grade - Maximum = 100

Instructor	Quarter	High School Physics					F	T	V#	Hypothesis
		Yes	No	Yes	No	S. D.				
		N <sub>1</sub>	N <sub>2</sub>	Mean						
Collins Kruglak & Valasek All three (C., K., V.)	Fall '51	137	17	58.7	56.9	13.4	14.4	1.20	.51	Accept
	Spring '52	155	17	57.1	49.2	11.8	10.6	1.17	2.64**	Reject
	F '51, Sp '52	292	34	57.8	53.0	12.6	13.2	1.10	2.08*	Reject

## Cooperative Physics Test for College Students

Maximum = 50

Collins Kruglak & Valasek All three	Fall '51	137	17	28.9	28.2	7.3	11.5	2.59**	.24	Accept
	Spring '52	155	17	27.3	20.1	8.1	5.1	2.33**	5.01**	Reject
	F '51, Sp '52	292	34	28.1	24.1	7.8	9.8	1.53*	2.21*	Reject

## Laboratory Performance Test

Maximum = 185

Collins Kruglak & Valasek All three	Fall '51	137	16	79.2	82.8	29.0	27.0	1.10	.47	Accept
	Spring '52	153	17	86.5	66.2	32.5	25.2	1.57	7.87**	Reject
	F '51, Sp '52	290	33	82.3	74.2	32.1	27.3	1.34	1.40	Accept

# Aspin-Welch Test for non-homogeneous variances

\*\* Significant at the one per cent level

\* Significant at the five per cent level

cont in our samples. Consequently, the data on the contribution of high school physics must be interpreted with great caution. The high correlation coefficients between a pre-test in high school physics and the theoretical achievement measures support the hypothesis that a high school physics course has probably a favorable effect on achievement in college physics as measured by the usual criteria. (See Table 2) On the other hand, from the low correlation coefficient between the same pre-test and the performance test a reasonable conclusion is that the influence of high school physics on laboratory performance test scores is not likely to be significant.

Additional evaluation will be needed to demonstrate conclusively the effect of high school physics on the theoretical and performance achievement in college physics.

#### Modification of Performance Tests

Experience with performance tests during 1951-52 led to several changes in the design, administration, and scoring of test items. More items with a high performance component were introduced. The credit allocation for the theoretical phase of each item was reduced or eliminated; thus no advantage accrued to groups taking the test at a later time because of information leakage on method. Each item was graded by one instructor for all the students in the course. One week before the test each student was given a sheet on which the test items were described in terms of the given apparatus and the problem to be solved or measurement to be made. Thus every student was given the same opportunity to make a theoretical preparation for the test. During the test all the needed formulas were either written on the board or were available on a mimeographed sheet.

The effect of the modified testing procedure was investigated during spring quarter 1952. A sample of 81 Institute of Technology students was drawn from physics

Table 2

Means, Standard Deviations, Reliabilities, and Correlation Coefficients of Initial and Final Measures

## University of Minnesota - Physics 7

Subsample 9<sup>#</sup>, N = 67 - Institute of Technology - Sophomore - Male - Minnesota High School Graduate - 1 year of High School Physics - Non-transfer students - Fall quarter 1951, Dr. T. Collins

Subsample 12, N = 81 - Same as subsample 9, but primarily third-quarter freshmen - Spring quarter 1952, Dr. J. Valasek

Variables	Max. Score	Mean	S. D.	X <sub>1</sub>	Correlation Coefficients			
					X <sub>2</sub>	X <sub>4</sub>	Y <sub>1</sub>	Y <sub>2</sub>
X <sub>1</sub> , High School Percentile Rank	(10p)	(6.27)	(1.39)					
	10p	6.02	.99					
X <sub>2</sub> , American Council on Education	(200)	(122.59)	(16.92)	(.37**)				
Psychological Examination-	200	125.86	19.00	.35**				
1947 Raw Score								
X <sub>4</sub> , Cooperative Physics Test	(78)	(24.78)	(9.94)	(.05)	(.45**)			
Revised Series Form S	78	27.64	12.48	.38**	.37**			
(High School)								
Y <sub>1</sub> , Cooperative Physics Test for	(50)	(29.57)	(7.60)	(.30*)	(.53**)	(.69**)		
College Students-50 selected	50	28.30	8.01	.47**	.44**	.72**		
items from forms E and F								
Y <sub>2</sub> , Final Grade in Physics 7	(100)	(59.21)	(13.42)	(.42**)	(.32**)	(.56**)	(.77**)	
	100	56.49	12.42	.40**	.35**	.62**	.74**	
Y <sub>3</sub> , Laboratory Performance Test	(300)	(130)	(35.90)	(.33**)	(.40**)	(.41**)	(.57**)	(.68**)
	185	82.26	46.41	.16	.16	.25*	.32**	.33**

# Values in parentheses

p - probits

\*\* Significant at the one per cent level

\* Significant at the five per cent level

7, taught by Prof. J. Valasek. This group was given substantially the same instructional exposure as the subjects on which the original findings were based (3). The means, standard deviations, and correlations coefficients for the two groups are shown in Table 2. It is clear that the means and the standard deviations on the initial and final measures do not differ greatly between the two samples. The only exception is the standard deviation on the performance test which increased by a factor of two for the spring quarter group.

The correlation coefficients between the initial measures, between the initial and final measures and the theoretical criteria did not differ substantially from the corresponding values for the fall quarter. However, the correlation coefficients between the spring quarter laboratory performance test and all the other measures have been decreased to about half of the corresponding coefficients for the fall quarter group. For instance, the correlation between the performance score and High School Percentile Rank has been reduced from .33, significant at one per cent, to .16, which is not significantly different from zero. The same is true of the correlation coefficient between performance scores and ACE. Apparently the aforementioned changes in the performance test account for the reduction in the size of the coefficients. The highest correlation is .33, between the performance test and the final grade in the course. Although the coefficient is significantly different from zero at the one per cent, its low value is indicative of the difference in the abilities and skills sampled by each of the two measures. In addition there is some intercorrelation between the two measures since the performance test accounts for about 7 per cent of the final grade. The low positive correlation is simply a measure of the overlap in terminology and apparatus between the laboratory work and lectures.

### Item Analysis

The Davis technique (1) was used to find the discrimination and difficulty indices of the performance test in Physics 7 for spring quarter 1952. The results are shown in Table 3. Four of the nine items were identical with those given during the fall quarter. The criteria used were the total score on the test and the final course grade. The average discrimination and difficulty index for the entire test was practically unchanged for the total score. The discrimination index was somewhat lower for the spring group with the final grade as a criterion. However, for the items common to both quarters the average discrimination and difficulty indices were higher on both criteria for the spring group. Hence the modifications in the performance test resulted in items that on the average discriminate better between the good and poor students and come closer to the optimum difficulty of 50 per cent.

### Hoyt Reliability

The Hoyt reliabilities (2) of the performance test were computed for the long items, short items and the total test. The values for the fall and spring quarters are shown in Table 4. It is seen that the reliabilities are highly comparable. The large difference in the reliabilities of the long and short item parts suggested an investigation of the test reliability as a function of the number of parts in a test.

During the fall quarter of 1952 the graders of the performance tests were asked to record the credit allocations for each part of a given item. For example, in one item the subject was asked to find the acceleration and the average velocity of a ball rolling down an inclined plane. The credit allocation for a perfect score was: 1 point for the measurement of the plane's length, 2 points for the time determination, 2 points for the correct acceleration value, and 2 points for the correct velocity value.

Table 3

## Item Analysis of the Laboratory Performance Test

University of Minnesota - Physics 7

Subsample 12 - Spring 1952 - SS 12

Subsample 9 - Fall 1951 (in parentheses)

Short item test Item No.	Indices*				Item Ave.	
	Discrimination Criteria**		Difficulty Criteria**		SS.12 N = 81	All Phys. 7 N = 221
	1	2	1	2		
M-138 m	38(35)	25(11)	46(31)	42(36)	35	33
M-106 m	56(38)	37(33)	40(38)	44(35)	31	29
M-71	27(14)	28(1)	25(31)	26(30)	12	12
M-36, 31b	43	10	28	29	15	13
M-222	13	17	53	53	53	51
H-2	12	14	59	59	63	63
E-93	11	10	55	54	63	65
S. Average	28	20	44	44	39	38
Long Item Test						
M-151	37(41)	10(12)	52(51)	47(51)	46	46
M-79 m	43	12	50	54	58	60
L. Average	40	11	51	51	52	55
S. & L. Average	31	18	45	45	45	45
Average of Common						
Items SS. 12 & 9	40(32)	25(14)	41(38)	40(38)		

\* Davis Indices

\*\* 1 = Total score on laboratory  
performance test  
2 = Final grade in physics 7

A sample of 62 Physics 7 students was drawn. The subjects had the same characteristics as those of the previous two quarters. The Hoyt reliability of the performance tests for this group was found in two ways: by using the total score on each item, and by repeating the calculations with part scores. In the first case there were 10 numbers for each individual grade; in the second trial there were 30 parts. The Hoyt reliability for the first procedure was .60, which was very nearly the same as that for the spring quarter group. On the other hand the coefficient went up to .75 when the subdivided parts were used. The process was repeated with two samples of students from a different course.

Two random samples of 62 and 64 male students were drawn from Physics 1a, fall quarter 1952. The Hoyt reliability was computed with 6 items of a performance test as well as with the 21 parts of the items. The coefficients were .44 and .43 for the first method and .89 and .93 for the second scheme as shown in Table 4. There appears to be some very strong evidence in support of the relationship between the Hoyt reliability and the number of parts or items in a performance test. The distribution of an item score into several subscores corresponds effectively to an increase in the number of test items. Thus, the mean square between individuals increases relative to the residual mean square with a corresponding increase in the reliability coefficient. When the scores on each item are distributed among the various tasks of the item, the reliability coefficient rises to an acceptable value. The problem of the low reliability of a performance test appears to have been solved satisfactorily. Another advantage of subdividing the score of an item into several components is the simplification of the grading procedure.

#### Implications for Further Research

The low correlation coefficients between the laboratory performance tests in Physics and other measures of aptitude and achievement suggest the need for in-

Table 4  
 Hoyt Reliability of Performance Tests as a Function of  
 the Number of Items - University of Minnesota

Course & Quarter	N	No. of Items	No. of Scores	Hoyt r
Physics 7, Fall 1951	67	3	3	.22
" " "	"	8	8	.59
" Spring 1952	81	2	2	.24
" " "	"	7	7	.60
" Fall 1952	62	10	10	.60
" " "	"	10	30	.75
Physics 1a, Fall 1952 Sample 1	62	6	6	.44
" " "	"	6	21	.89
Physics 1a, Fall 1952 Sample 2	64	6	6	.43
" " "	"	6	21	.93

vestigating the relationship between various measures of laboratory work.

A number of pencil-paper tests on laboratory work in elementary physics have been developed. Several of the tests have been administered during 1952-1953 to students enrolled in three different physics courses at the University of Minnesota. The results are being analyzed and the relationships between the pencil-paper and performance test scores are under investigation.

#### Summary and Conclusions

The investigation dealt with the relationship between performance tests in physics and other measures of scholastic aptitude and achievement. Students with and

without high school physics were compared on theoretical and performance tests in college physics. The dependence of the Hoyt test reliability on the number of items and the number of sub-scores was explored.

The major findings of the study were:

1. Students with high school physics background appeared to be superior to those without high school physics on conventional measures of achievement in college physics. However, the differences on the performance tests in the laboratory were probably not significant.
2. A reduction of the theoretical component and other modifications of a performance test result in a lower correlation coefficient between it and other measures of scholastic aptitude and achievement in physics.
  - a. The correlation coefficient between the first-quarter performance test in physics and ACE was .16, not significantly different from zero.
  - b. The correlation coefficient between the performance test and High School Rank was also .16.
  - c. The correlation coefficient between the performance test and the final grade in physics was .33, significant at one percent.
3. The various changes in the administration of the performance tests resulted in small but desirable changes in the average discrimination and difficulty indices.
4. The Hoyt reliability depended on the number of items in a test. It was found that the reliability of a 6-item performance test could be raised from .43 to .93 by using 21 subscores.

Paper-pencil tests on laboratory work in physics have been constructed and a number of them have been administered. Their relation to performance tests is being investigated.

## References

1. Davis, F. B. Item Analysis Data. Cambridge: Graduate School of Education, Harvard University, 1949.
2. Hoyt, C. J. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, 153-160.
3. Kruglak, H. Performance Tests in Physics. Technical Report No. 1, ONR Project NR153-148. Minneapolis: Physics Department, University of Minnesota, October 1952
4. Wall, C. N., Kruglak H., & Trainor, L. E. H. Laboratory performance tests at the University of Minnesota. Am. J. Phys., 1951, 19, 546-555.